Caching in Multidimensional Databases

István Szépkúti

One utilisation of multidimensional databases is the field of On-line Analytical Processing (OLAP). The applications in this area are designed to make the analysis of shared multidimensional information fast.

On one hand, speed can be achieved by specially devised data structures and algorithms. On the other hand, the process of analysis is a cyclic one. In other words, the user of the OLAP application runs his or her queries in a row. The output of the latest enquiry may be there (at least partly) in one of the previous results. Therefore caching also plays an important role in the operation of these systems.

However, caching in itself may not be enough to ensure acceptable performance. Size does matter: The more memory is available, the larger we gain by loading and keeping information in there.

Oftentimes, the cache size is fixed. This limits the performance of the multidimensional database, as well, unless we try and condense the data in order to move a greater proportion of them into the memory. Caching combined with proper compression methods promise further performance improvements.

In this paper, we investigate how caching influences the rapidity of OLAP systems. Different physical representations (multidimensional and table-based) are evaluated. For the thorough comparison, a model is proposed. We draw conclusions based on this model, and these conclusions are verified with empirical data. In particular, using benchmark databases, we show examples when one physical representation is more beneficial than the alternative one and vice versa.

Keywords: *compression, caching, multidimensional database, On-line Analytical Processing, OLAP*